大数据视阈中的明清进士家族研究
——以 CBDB、中华寻根网为例

刘 京 臣
(中国社会科学院 文学研究所, 北京 100732)

摘　要: 作为关系型数据库,“中国历代人物传记数据库”(China Biographical Database Project,简称 CBDB)因拥有 42 万多条中国历代人物传记数据而成为从大数据视角整体关照明清进士家族的首选。它的“亲属关系”查询,能够呈现亲属关系网络,方便进行进士家族的梳理。同时,也应承认 CBDB 存在着匹配不准确、字段多重复等问题,这些问题可以借助家谱等文献得以解决。现阶段,中华寻根网提供了 2392 种家谱影像全文,便于我们利用爬虫程序提取家谱目录中的相关信息,进而自动判断进士家族,这是对 CBDB 的有益补充。在此基础上,借助社会网络分析,可以直观呈现以血缘、姻亲为核心的进士家族世系图谱。姻亲关系能够将多个进士家族之间联系起来,可以使视角由单个进士家族向有关联的多个进士家族转移,例如可以考察这些姻亲家族在治经、传经等方面是否存在影响等。依赖家谱留存下来的家族成员的文学作品,既是考察家族文学发展流变的基础,又是建构更加广泛的社会网络分析图谱的基石。

关键词: 大数据;明清进士家族;CBDB;中华寻根网;家谱;社会网络分析
中图分类号: K 23　　**文献标识码:** A　　**文章编号:** 1000-5919(2019)04-0096-13

科举制是我国历史上一项非常重要的人才选拔制度。进士科本为诸科之一,后来逐渐成为最重要的一科。姚合说“蹇钝无大计,酷嗜进士名”①,《唐摭言》也说:“缙绅虽位极人臣,不由进士者,终不为美,以至岁贡常不减八九百人。其推重谓之‘白衣公卿’,又曰‘一品白衫’;其艰难谓之‘三十老明经,五十少进士’。”②可见时人对进士之推重。

学界对进士群体较为关注,取得了一系列成果。这些成果中,有一些是在历代登科录、题名录、碑录以及方志等基础上选取区域进士群体为考察对象的③,极大地推动了区域文化与进士群体研究;还有一

收稿日期: 2018-11-30
作者简介: 刘京臣,男,山东莱州人,中国社会科学院文学研究所副研究员。
基金项目: 本文为国家社会科学基金重点项目“元代文学地图数字分析平台”(项目编号:18AZW008)、国家社会科学基金重大项目“基于大数据技术的古代文学经典文本分析与研究”(项目编号:18ZDA238)阶段性成果。

① 姚合:《寄陕府内兄郭冏端公》,《全唐诗》卷四九七,北京:中华书局 1960 年版,第 15 册,第 5647 页。
② 王定保:《唐摭言·散序进士》,上海:上海古籍出版社 1978 年版,第 4 页。
③ 例如多洛肯《明代福建进士研究》(上海:上海辞书出版社 2004 年版)、《明代浙江进士研究》(上海:上海古籍出版社 2004 年版)、《清代浙江进士群体研究》(北京:中国社会科学出版社 2010 年版),李润强《清代进士群体与学术文化》(北京:中国社会科学出版社 2007 年版),朱东根《海南历代进士研究》(海口:海南出版社、南方出版社 2008 年版),徐红《北宋初期进士研究》(北京:人民出版社 2009 年版),丁辉、陈心蓉《嘉兴历代进士研究》(合肥:黄山书社 2012 年版)与《明清嘉兴科举家族姻亲谱系整理与研究》(北京:中国社会科学出版社 2016 年版),陈尚敏《清代甘肃进士研究》(兰州:甘肃人民出版社 2013 年版),毛晓阳《清代江西进士丛考》(南昌:江西高校出版社 2014 年版),严其林《镇江进士研究》(上海:复旦大学出版社 2014 年版),邱进春《明代江西进士考证》(北京:中国社会科学出版社 2015 年版)等著作。

些则在此基础上进一步考察进士家族①,探讨这些家族的形成、分布、流动、姻亲以及对地方的影响等。这些研究之所以能取得令人瞩目的成就,绝大部分如郭培贵在梳理明代进士家族时所说的:“对所见明代各科《进士登科录》《进士履历便览》《皇明三元考》《献征录》《类姓登科考》及清官修畿辅和各省《通志》中的《选举志》与人物传记等文献中的相关史料进行了全面搜集、整理与考证。”基于对上述文献的一一爬梳,郭培贵认为明代“273 年间共产生 2088 个进士家族,与其相关的进士总计为 4970 名,占明代进士总数的 20.21%”②。正因为有传统文献梳理为基础,得出来的结论才较为令人信服。

若从另外的角度看,我们如何检验学者们关于进士家族确立③、进士家族数量的判断?或者简单地说,某时段、某区域有多少进士家族,能否比较迅速地探究清楚?构成进士家族的进士们,在家族中分属哪些世系、支派?他们的姻亲状况如何?联姻的家族如果也是进士家族,那么不同的家族之间是否会在治经、传经等问题上产生影响?能否从家谱和家族成员的作品中,细化姻亲、交游、师友等关系?能否借此挖掘出构建更为全面的家族世系图谱的有效数据?借助技术手段,或可回答上述问题,我们以 CBDB 和中华寻根网为例进行尝试。

一、CBDB:明清进士家族的初步梳理

作为关系型数据库的 CBDB,以宋元明人传记资料索引及《清代人物生卒年表》,正史列传、墓志铭、墓表、地方志列传等史料,文集之祭文、序、记、书信等作品以及郡守年表、会要、实录等官方文书为主要资料来源,同时还与“明清档案人名权威数据查询”、麦吉尔大学“明清妇女著作”数据库以及京都大学“唐代人物知识库”等开展合作,截至 2019 年 4 月,单机版共收录中国历代人物传记资料 422607 条④。

它以姓、名、生卒、地址、别名、著述、职官、入仕、事件、社会区分、亲属关系、社会关系、财产、出处、社交机构为字段,尽可能将入库的人物信息细化,并在诸多子库之间建立起关联。在 CBDB 看来,每一位被纳入数据库中的个体,都“被视为被关系网络界定且能够被量化和分析的实体”,基于这种理念,拥有 42 万多条数据的 CBDB 可以将某个时段某个区域内的所有个体以某种或某些特定方式呈现出来——当然,呈现的背后是量化与分析。

且看如何利用 CBDB 介入到明清进士家族研究领域。从理论上讲,该数据库能将明清时期整个版图内的进士纳入考察视野,通过亲属关系等判断进士之间的关系,从而推断他们是否来自同一家族。为方便论述,且以明清时期的萧山为例。据 CBDB 可知明清时期(1368—1911),萧山⑤共有进士 142 人,明代 40 人、清代 102 人。

从 CBDB 的角度来考察明清两代的萧山进士群体能否归并到相关家族中,这 128 位进士是最直接的研究对象。当然,我们也可以从登科录、方志、朱卷集成等文献中一一爬梳,但那样的操作在当下失去

① 如张杰《清代东北边疆地区的科举进士家族》(《中国边疆史地研究》2000 年第 3 期)、张杰《清代科举家族》(北京:社会科学文献出版社 2003 年版)、叶可汗《明代福建进士家族研究》(辽宁师范大学 2012 年硕士学位论文)、陈秋露《明代江西进士家族研究》(江西师范大学 2014 年硕士学位论文)、郭培贵《明代进士家族相关问题考论》(《求是学刊》2015 年第 6 期)以及夏汉宁、黎清《文化地理视域下的宋代江西籍进士家族》(《江西社会科学》2017 年第 11 期)等皆为此类。

② 郭培贵:《明代进士家族相关问题考论》,《求是学刊》2015 年第 6 期,第 145 页。

③ 关于“进士家族”的标准,学界并无定论。郭培贵在《明代进士家族相关问题考论》中提出“‘五代直系亲属内有两名以上进士的家族’即为进士家族,也即五代之内不一定每代都出进士,但至少应有两名以上进士”(《求是学刊》2015 年第 6 期,第 144 页),本文采用郭先生之说。

④ 2019 年 4 月 15 日下载链接为:https://hu-my.sharepoint.com/:u:/g/personal/hongsuwang_fas_harvard_edu/EXe4gcr6HalEuoyIdRmOBXsBKKFUnEYYXnchcvl0JMEvjg?e=Oop9fa。

⑤ CBDB 仅收录乡贯为萧山,未包含户籍为萧山的进士,特此说明。

了技术的意义,借助于 CBDB 所提供的一系列查询与呈现方式,我们可以较为便捷地考察出这些进士们有无亲属关系。如果存在亲属关系,再利用社会网络分析软件将这些关系呈现出来。

通过分析,我们发现 40 位明代萧山进士形成了以张试、张谊、来天球、来汝贤、来经济、来三聘、翁五伦等为中心的 27 个独立原型,还有 13 人的亲属关系 CBDB 尚未梳理。这 27 个独立原型中,有 4 个以来氏为中心,分别是来天球、来汝贤、来经济与来三聘,以下是依据 CBDB 所呈现出来的来天球的亲属关系(图 1):

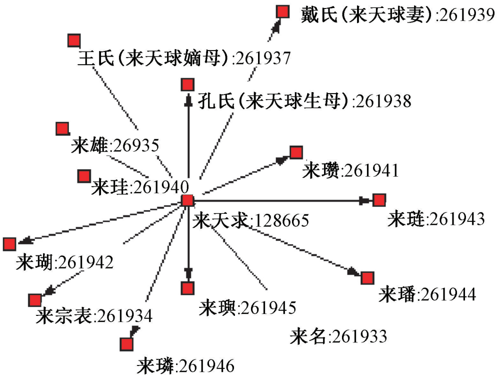

图 1

来天球的 13 种亲属关系,包含了其曾祖来名、祖父来宗表、父亲来雄,其兄珪、瓒、瑚、琏、璠、玙,其弟璘,嫡母王氏、生母孔氏,其妻戴氏等,未见与其他来姓进士有关联。由来汝贤的亲属关系来看,也未见与其他三位来姓进士有关联。来三聘的亲属关系显示他与来汝贤(ID224766)、来经济(ID224767)有关系,但这二人 ID 都与具有进士身份的来汝贤(ID128666)、来经济(ID205509)不相符,这证明 CBDB 中至少有两位来汝贤①、两位来经济,且与来三聘有关系的二位皆非进士。在来经济的亲属关系网络中出现的来天球(ID334474)亦非进士来天球(ID128665)。

借助于 CBDB 及其提供的一系列工具,我们可以较便捷地知道某个时段某地所有进士,还可以探究这些出自同一地域的进士群体的亲属关系(当然如果需要还能探究社会关系等),从理论上讲,这为构建进士家族创造了条件。但在尝试过程中,至少明代萧山一地的四位来姓进士,并没有显示出亲属关系,这意味着他们不能被纳入同一个家族中去。

明代如此,清代呢？据 CBDB 统计出来的清代萧山 102 位进士中,来姓进士 12 人,当我们以此为例进行考察时,发现通过 CBDB 的梳理,他们之间也不存在亲属关系,甚至连社会关系也不存在。这颇令我们怀疑:难道明清两代的萧山来姓进士之间竟然不存在亲属关系？至少目前从 CBDB 呈现出来的数据表明,他们之间并不存在关联,更不用说出自同一个家族。

为什么会出现这种情况？我们推测可能有如下几种原因:

第一,明清两代的萧山来姓进士确非出自同一家族,遂无亲属关系。依常理,这个判断成立的可能性不大。

第二,由于数据不足,部分进士未被统计进 CBDB,从而导致一些原本可能建立起来的联系因这些关键人物的缺失而中断。

第三,由于算法原因,一些进士虽被 CBDB 收录,却因户籍类别问题而被排除在统计范围之外。如鲁琛、来宗道、陈伯龙、来斯行、来方炜、来集之等虽被 CBDB 收录,却皆未呈现在萧山进士群体中。

被 CBDB 认可为萧山进士的来天球、来汝贤、来三聘与来经济四人,皆为灶籍;虽被收录却未被认可

① CBDB 中共有三位来汝贤,ID 分别为 128666、224766、334480。

为萧山进士的来宗道、来斯行、来方炜三人也是灶籍,这两批人皆为绍兴府萧山县的灶籍,CBDB 出现了厚此薄彼的现象,这给我们带来了新的挑战:如果要依托 CBDB 进行较大规模的数据分析,如何检验数据? 就本文而言,我们要以 CBDB 为数据源探讨明清两代进士家族,首要问题是利用大数据梳理出某地所有进士,其次利用 CDBD 的"亲属关系"功能将这些被确认出来的进士们归并到各自的独立原型中,再来判断是否出自同一家族。目前来看,虽然 CBDB 确认进士身份的准确率较高,但是一旦与某地结合起来、确认该地的进士时,往往会出现滥入、失收等问题。就 CBDB 的数据来源而言,如果能将数字化的家谱文献纳入考察视野,与现有的数据相匹配,能在一定程度上解决滥入特别是失收的问题。

第四,人物重名是很常见的现象,CBDB 同样要面临和解决这个问题。在无法确知重名者为同一人的情况下,各自赋予唯一 ID 进行区分确为明智之举。本文中涉及的来天球便有两个 ID(128665、334474),在 CBDB 看来,这便是有两位来天球。实际上这两个 ID 指向同一人,即 ID 为 128665 的进士来天球。与之相似的,还有来汝贤、来经济。正因为 CBDB 将一人认定为两人或多人,且赋予每人唯一 ID 的做法,使得它虽然能勾勒和展现一些亲属关系,却往往伴随着疏误出现——它中断了本应建立起来的关系。除了来氏外,上文统计出来的明代进士张试、张谊兄弟也存在同样问题。

通过重新修正来天球等人的 ID,确定惟一身份之后,便会发现明代萧山的四位来姓进士之间确实存在亲属关系。如何对重复数据进行清洗,如何确定亲属关系? 除了进一步完善规则进行数据挖掘外,可能还需要将朱卷、登科录、碑录、《类姓登科考》、方志中的选举志等文献中关于进士的信息尽可能多地纳入数据库中,以便进行数据分析与数据挖掘。当然,这方面最直接的文献,首推家谱。家谱中对于世系、支派、亲属关系的记载,最为详实①。按照传统做法,从家谱入手,首先应当查询家谱目录,然后去庋藏地梳理相应的谱牒。在大数据时代,是否有更为便捷的途径? 答案是肯定的。

二、中华寻根网:以家谱为依托的社会网络分析

近年来,各界对家谱数字化的重视程度越来越高。爱如生开发的《中国谱牒库》专门收录历代谱牒类文献,据称达五万余种,是商用数据库的集大成者。还有一些机构——如中国国家图书馆与旗下中华寻根网②,上海图书馆与旗下上图家谱网以及美国"犹他家谱学会"图书馆网站等,以公益性质数字化了大量家谱。还有一些网站提供家谱信息查询,比较便捷的有上海图书馆家谱数据库、浙江图书馆家谱数据库以及台北"故宫博物院"家族谱牒文献资料库等。

前述拥有家谱全文影像的国家图书馆、中华寻根网、上海图书馆、上图家谱网以及"犹他家谱学会"图书馆网站等,都利用技术手段对源文件进行了处理。国内网站多采用 Flash 或 Pdf 封装法,需调用 Flash 播放器或 JAVA 在浏览器中打开相应文件;"犹他家谱学会"图书馆网站虽然可以直接查看、下载 JPG 源文件,却采用了动态变换源图地址以及切分原图等技术。

这些数据源中,只有中华寻根网提供了家谱目录,仅这一点就使其意义远超其他数据源。于是,我们将关注点转到中华寻根网所提供的 2392 种家谱全文影像上。首先抓取 Flash 阅读器文件进行反编译,分析其与后台动态接口交互的参数,再通过 Python 编写爬虫程序,批量抓取分析目录数据。在这 2392 种提供全文影像的家谱中,凡是目录中有"选举""科第""科名""科甲""题名""文科志""科宦""职官""仕宦""绅宦""宦达""仕籍""仕履""乡会榜"等,该家族拥有进士、举人、贡生等功名人士的可

---

① 当然,家谱也可能存在着"出于闾巷,家自为说,事非经典,苟引先贤,妄相假托,无所取信"(颜师古语),甚或冒认始祖、虚美隐恶等问题,所以在使用家谱时,一定要与其他相关文献互相印证,关于这点张剑《宋代以降家族文学研究的理论、方法及文献问题》(《文学评论》2010 年第 4 期)所论甚为详备,可参见。

② 2018 年 6 月八日访问链接如下:http://www.ouroots.com.cn/index.jsp。2019 年 4 月 15 日访问提示连接超时,据中国国家图书馆·中国国家数字图书馆(www.nlc.cn/index_zt_3339.html)显示,中华寻根网正处于维护中。

能性越大①。

明清两代萧山的来姓进士,是否出自同一家族;如果是,他们之间存在怎样的亲属关系?

中华寻根网收录光绪二十六年(1900)、民国十一年(1922)两种《萧山来氏家谱》,前者卷七有"文武乡会榜"、后者卷八有"绅宦录",都将该家族的进士、举人等一一呈现出来。现在看似有了该家族的进士名单,但是还要去伪存真、辨别支派。"去伪存真",就是要判断进士身份。如果不是,则要剔除;如果是,就要"辨别支派",考察是否出自本家族,是何世系支派,与其他进士是何关系等。

据家谱可知,来廷绍为萧山来氏始迁祖,"以进士历官朝散郎、直龙图阁学士,兼运使加宣奉大夫。嘉泰二年出知绍兴府,未抵任,道卒。萧山子孙因家焉,萧山之有来氏,盖始此"②,并将他列为萧山进士之首③,称其为"宋绍熙癸丑进士"④,家谱中所列的明清两代萧山一地的其他 20 位来姓进士皆为其后代,可谓盛矣,兹列表(表 1)如下:

表 1

	10 世	13 世	14 世	16 世	17 世	18 世	19 世	21 世	22 世	23 世	24 世
大支				来方炜	来垣	来谦鸣				来珩	来熊
二支						来珏		来宗敏	来学醇		
四支			来斯行	来集之	来燕雯		来起峻	来煦	来凤郊		
五支	来天球	来汝贤 来经济	来宗道			来益清					
六支		来三聘									

表 1 所呈现的是家谱中收录的 20 位来姓进士,明代有来天球、来汝贤、来经济、来三聘、来宗道、来斯行、来方炜与来集之 8 人,余下 12 位乃是清代进士。据《明清进士题名碑录索引》,明清两代萧山一地共有 20 位来姓进士,数量与家谱所载吻合,但有两处不同:

第一处不同是,《明清历科进士题目碑录》收录了康熙三十九年(1700)两位来姓进士:三甲 130 名来燕雯与三甲 235 名来楫,称此二人皆为"浙江绍兴府萧山县人"⑤。来燕雯出自萧山来氏家族,是四支第 17 世,关于他的记载,碑录、方志与家谱等几种文献吻合。《浙江通志》称来楫为萧山人,然而光绪、民国两种家谱中,皆未收录来楫。

第一,家谱凡例称:"子姓徙居他处,或出仕在外者,今皆邮寄书函,或遣人采访,故谱中较前详备,以示无遗。凡远如滇、黔、闽、粤,近如本省乡僻,寄递谱局稍迟者,则入补遗,以备将来增入。"⑥可见凡本家族子孙徙居他处或出仕在外,无论远近,皆力求搜访,载之谱牒。光绪、民国两种家谱除收录萧山来姓进士外,还都有"广世系会榜",连陕西三原人来聘、来复都一并收录,如果来楫果真出自萧山来氏家族,焉有不收之理?

第二,已知来燕雯出自萧山来氏家族,如果来楫果如碑录、《浙江通志》所言亦为萧山来氏,且此二人为同榜进士,家谱中为何只字不提?《清代科举人物家传资料索引》《清代朱卷集成》同样未收录来楫。

① 当然,这既不意味着目录中没有呈现上述标识的家族定然没有进士、举人等,也不意味着凡出现上述标识的家族中定然就有进士产生,例如民国二十四年《鄞城施氏家乘》卷六"仕履表"仅收录该家族的举人、贡监、生员等,这说明该家族并无进士。

② 《萧山来氏家谱》卷首《原姓氏支派》,民国十一年(1922)刻本。

③ 龚延明《宋代登科总录》未收录来廷绍,此公是否为绍熙癸丑进士,待考。

④ 《萧山来氏家谱》卷八《绅宦录·会榜》。

⑤ 《明清历科进士题名碑录》,台北:华文书局 1969 年版,第 3 册,第 1694 页。

⑥ 《萧山来氏家谱》卷首《雪珊公续修宗谱凡例》。

第三,来楫有无可能本为他姓而冒姓来氏?《绍兴府志》在康熙三十九年进士中收录了来楫,称:"《县志》本姓陈,知县,俱萧山人。"①复核《萧山县志稿》,康熙三十八年举人中有来楫,"本姓陈,萧山籍,山阴人";三十九年进士中称其为"吏部主事,《绍兴府志》作知县"②,据此推知来楫本姓陈,来氏家谱定然不会收录此人。

第二处不同是,《明清历科进士题目碑录》称来珏为"浙江杭州府仁和县人"③,家谱却收录其中。《浙江通志》言及康熙四十一年(1702)举人时称来珏为"仁和人,丙戌进士"④,在言及康熙四十五年(1706)进士时,又称其为"萧山人,知县"⑤;《绍兴府志》则称来珏"《碑录》仁和,《旧志》萧山人,知县"⑥。那么此人到底是仁和人,还是萧山人?《清代科举人物家传资料索引》《清代朱卷集成》亦未收录,我们从家谱中寻找答案。

民国谱称来珏:"尔绳子,行彝一,字子苍,号玉峰。生顺治戊戌十二月初四日。仁和廪生。康熙壬午举人,丙戌登施云锦榜进士,进阶文林郎,福建福州府永福县知县。"⑦来珏之父来尔绳为"钱塘廪生,康熙丁亥科恩贡"⑧,可见最迟在其父时即亦迁出萧山。来氏家谱"凡例"称"子孙徙居他处或出赘别乡,并注于世系表内,使后世知所考焉"⑨,所以从家族角度看,将来珏纳入家谱之中无可厚非;从地域角度看,自非萧山进士。

可见,无论是从收录进士数量,还是从人物关系的准确呈现等方面,家谱都优于现阶段的 CBDB。下面我们依据家谱呈现萧山来氏第五支的社会网络图谱,该支在明清两代共有 5 位进士,数量占整个家族进士数量的四分之一,还产生了该家族在明代最早的一位进士来天球。呈现这个分支,有典型性。

如前所述,来廷绍是萧山来氏始祖,二世来师安,三世来大德,四世来贤,五世来尚宾,六世来韶,七世来思名。思名居长河,即是萧山来氏六支之祖。有子二,长励、次仪。来励三子,长阜、次浚、三川,分别是大支、二支与三支之祖;来仪亦三子,长英、次雍、三岊,分别是四支、五支与六支之祖。兹将该家族第五支,以来天球、来汝贤、来经济、来宗道、来益清五位进士为中心的世系图(图 2)呈现如下⑩:

图 1 是据 CBDB 所呈现出来的来天球亲属关系,包括其曾祖、祖、父、兄弟、嫡母、生母与其妻在内共 13 人。CBDB 数据库中,无论这些人物之间是否存在代际关系,凡有联系,皆是直接关联,所以图示呈现出明显的扁平化态势,仅借助它无法直接判断二者之间的亲疏远近和代际关系。

相较之下,世系图侧重呈现血缘关系与婚姻关系,图 2 起于始迁祖来廷绍,终于来益清之子来慎修,历 19 世、共 114 人、39 个家庭。其中的三角顶点代表男性、圆形顶点代表女性,他们之间无方向的蓝色实线代表婚姻关系,弧(即有方向的箭头)代表父母与子女之间的关系。它既强调人与人之间的直接联系,例如父子、夫妻;又重视代际之间的关系,如曾祖与曾孙等,所以世系图的呈现会更加直观。另外,世系图还可以直接计算两人之间的距离远近,判断两人与远祖或世系图上任意一人的关系。

世系图虽然能够呈现血缘、婚姻关系,但图 2 的主体部分,或者说以中国古代家谱为代表的世系图

---

① 《绍兴府志》卷三一《选举志二·进士》,乾隆五十七年(1792)刻本,第 60 页。
② 《萧山县志稿》卷一三《选举表》,民国二十四年(1935)刻本,第 50 页。
③ 《明清历科进士题名碑录》,第 3 册,第 1716 页。
④ 《浙江通志》卷一四四《选举二十二》,乾隆元年(1736)刻本,第 13 页。
⑤ 《浙江通志》卷一四二《选举二十》,第 33 页。
⑥ 《绍兴府志》卷三一《选举志二·进士》,第 61 页。
⑦ 《萧山来氏家谱》卷二。
⑧ 《萧山来氏家谱》卷二。
⑨ 《萧山来氏家谱》卷首《后江公重修族谱凡例》。
⑩ 凡世系图中的进士,皆在姓名之后加标★符号,特此说明。

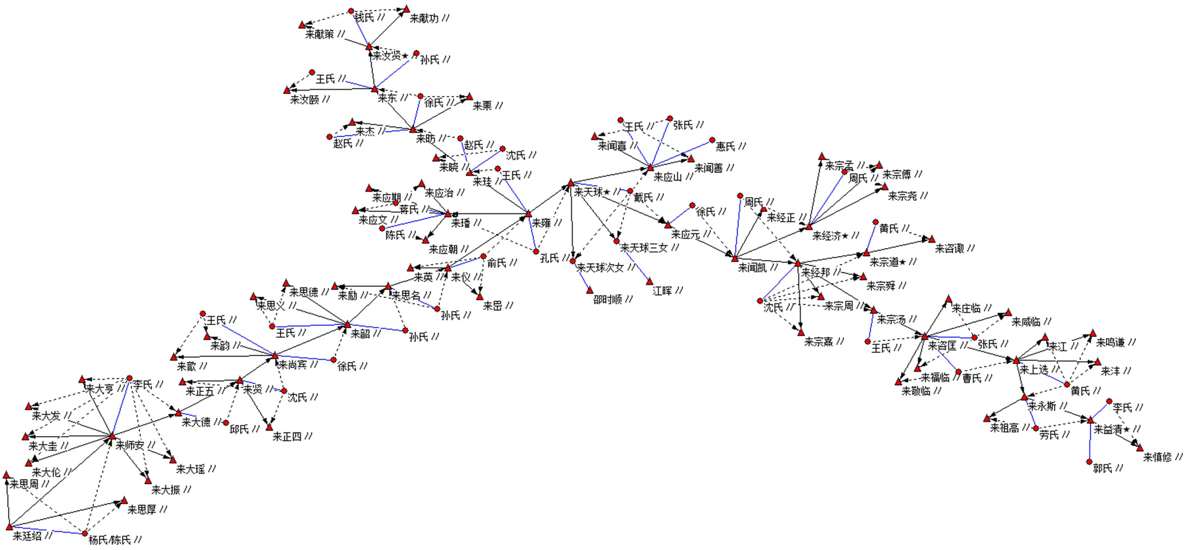

图 2

的主体泰半是血缘关系的展现,是否可以利用家族之间的联姻来更多展现婚姻关系?答案也是肯定的。我们以来天球为例,进一步细化他的世系(图 3):

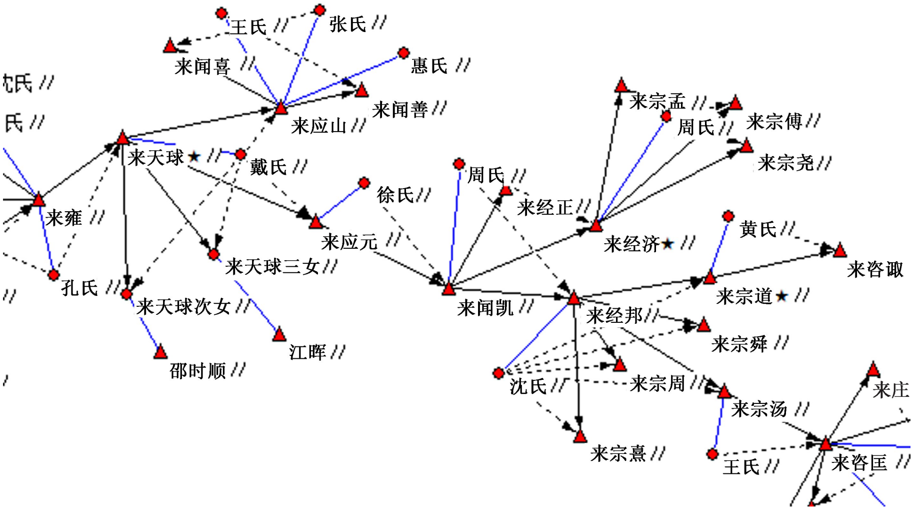

图 3

其一,据 CBDB 可知,来天球有兄 6 人,弟 1 人,是亲兄弟,还是堂兄弟?如果是堂兄弟,各出自哪一支?皆不得而知。据家谱,来珪为天球同父异母(王氏)之兄,来播与来天球皆为孔氏所生;来瓒、来琏皆四支祖来英之子,来瑚、来玙、来璘皆六支祖来岊之子,这 5 人都是来天球的堂兄弟。如果将整个家族六支的世系图呈现出来的话,那么他们兄弟之间的关系会更加直观。

其二,据登科录,来天球曾祖来名、祖父来宗表、父亲来雄。家谱则称其曾祖为来思名,“行一,字以实,一字叔顾,号潭居”;祖父为来仪,“行三,字宗表,号冠山”;父亲为来雍,“行五,字淳熙,号侃如”。故而推测来天球曾祖当为来思名而非来名,登科录疑夺字,其祖父即来仪,字宗表,登科录以字行。至于其父,雍、雄音近,似当从家谱作雍。

其三,来雍与孔氏生天球,天球与戴氏有二子、四女。图 3 呈现的虽侧重于父系谱系,却可以利用婚

姻关系,即通过来氏与其他家族的联姻,将不同的家族关联起来。《后江公族谱凡例》在论及家族成员的婚嫁时说:“妻室不书名,以其不外见也。书曰‘娶某乡某氏’,曰‘继某乡某氏’。或其祖、父有官者,而书曰‘某官之女’,重其妻与己齐也,且识外家也。”“所生之女,书曰‘女适某处某人’。”①这种凡例便于我们从家谱中发现更多的与姻亲家族相关的信息。邵蕃称,来天球“子二,长应元,廪生,娶钱塘宪副一轩徐公潭女;次即应山,娶仁和宪副北屏惠公隆女,继都宪张公本孙女。女四,一适山阴周义封天锡公子典膳周淙;二适予仲子临淄知县时顺;三适仁和江尚书文澜公子编修晖;四适钱塘张腾霄”②,这也与家谱记载相吻合。

徐潭、惠隆,弘治六年(1493)同榜进士,一为三甲 194 名、一为三甲 195 名,其女分适来天球长子应元、次子应山;应山继妻,乃钱塘张本之孙女,张本是成化十一年(1475)进士。来天球二子所娶之妻皆来自“进士之家”,其女所适之邵氏、江氏则为“进士家族”。次女适邵时顺③,余姚邵氏是科举世家,人才辈出,仅明代就有 17 位进士,这其中就包括邵时顺之父成化二十年(1484)进士邵蕃、邵时顺之子嘉靖二十三年(1544)进士邵漳,邵漳也就是来天球的外孙。来天球三女适仁和江晖,江晖祖父江玭是景泰二年(1451)进士,其父江澜是成化十四年(1478)进士,其从子江圻、江圻之子江铎分别是隆庆二年(1568)、万历二年(1574)进士,故而江晖亦出身进士家族。通过来天球子女的婚嫁,可见来氏与邵氏、江氏、惠氏、徐氏以及张氏五大家族之间有着较为密切的联系(图 4):

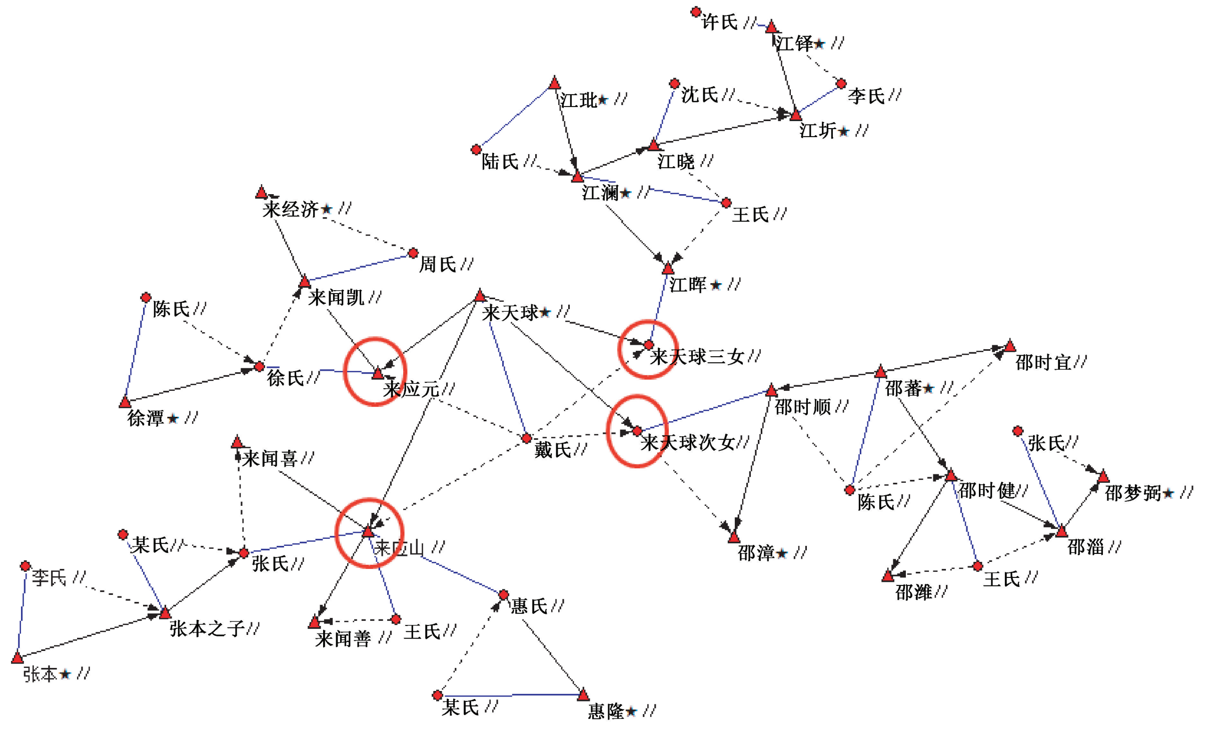

图 4

---

① 《萧山来氏家谱》卷首《后江公重修族谱凡例》。

② 邵蕃《明任中宪大夫陕西按察使致仕奉诏进阶荣禄大夫两山来公墓志铭》,《萧山来氏家谱》卷五《赠言》。

③ 邵时顺“以例贡官山东莱州府滨州州判,调山东青州府莒州州判,三载考最,敕授征仕郎,升青州府临淄县知县……配来氏,敕封孺人,加封安人,生一子漳”(《余姚邵氏宗谱》卷一,光绪十四年刻本),正与来氏家谱所称之“临淄知县”相吻合。

来天球二子、二女将萧山来氏与余姚、仁和、钱塘三地的五大家族联系了起来,以上仅仅是涉及六个家族的五段姻亲,如果能将更多的与他们有姻亲关系的家族关联起来,相信将会更有助于我们理解家族与家族、与地域、与姻亲、与教育、与经学、与文学等领域的关系。

其四,除了依靠血缘、姻亲关系计算和分析家谱世系外,还可以借助于其他社会交往活动来建立不同家族之间的联系,进而绘制新的图谱①。民国谱有《赠言》两卷,对家族中"其人为《乘》所不及载者,或《乘》已载其人而行事又别有所表见者"收录颇多,收录的标准是"诗章、寿章之类无关行事者一概不录,而博收之中仍有谨严者存焉",目的有二:一是"使后人读之而有所兴起,以仰承之孝友之心",二是"补二《乘》之阙"②。《赠言》这种强调记录行事的做法,可以为我们梳理传主生平事迹提供更为详实的资料。这是从家族与传主的视角着眼,如果我们将视角转向撰写者呢?来氏家族的征言者,会向具有怎样身份的族外人士征求像赞、传略、墓志等?撰者与征言者、传主与征言者是什么关系等,都可以成为社会网络分析的重要材料。

《赠言》所示,为来氏作像赞、传略、墓志等的族外人士,宋代有辛弃疾、徐理等人。来大章认为辛弃疾为来廷绍所作的《宋宣奉大夫知绍兴府事来公墓志铭》"或经后人改篡,以致参差抵牾如此也",似不足信;徐理,绍兴府萧山人,宝祐四年(1256)文天祥榜进士。来师安,生于孝宗淳熙八年(1181)七月,卒于宝祐元年(1253)十月。理论上讲,徐理有为来师安作墓志铭的可能性,但其墓志不存,仅能于族藏抄本中得之,并不著名,"短章寂寥,殊不类传体,且文亦不雅驯",所以来集之怀疑"其为后人失志而补作无疑"。故而所谓辛弃疾、徐理所作墓志似乎皆不足信。

当然,《赠言》之中也有一些确凿无疑的记载。如前所述,邵蕃与来天球是儿女亲家,为来天球夫妇所作的墓志铭记载行事颇为详实可靠。还有一位是陶望龄,《赠言》卷二有他为来经济夫妇所作的《明贵州参议继山来公暨配周宜人墓志铭》,《歇庵集》中也收有此篇③。陶望龄曾为经济之弟经邦作《待赠冠岩来公行状》④,记载经济抚略四方、经邦操持家业之事甚详,惜家谱未收;还曾为来弘辉作过《孝义来功伯墓表》⑤,茅瓒为其作《明授引礼舍人南庄来公墓志铭》,家谱收录茅文而未收陶文。

除此之外,《赠言》中还收录有戴子静⑥、韩守正、魏骥、钱溥、何舜宾、钱福、张嵿、高台、翁五伦、沈束、朱赓、潘士藻、钱士鳌等人为家族成员所作的墓志。上举诸公,皆为进士,除潘士藻外都是浙江人士。如果这些墓志皆出自诸公手笔⑦,那么说明来氏家族与江浙一地的诸科进士们有着非常密切的联系,这类社会交往活动正是构建社会网络的重要节点之一。

三、进士家族与经学、文学之关系

以上我们关注了以血缘、姻亲为代表的世系图谱,并通过家谱中的像赞、传略、墓志等的写作,将家族成员与其他相关人物关联起来,为进一步丰富世系图,绘制更为宏观的社会网络分析图谱奠定了基

① 这是所有关系型数据库的题中应有之意,CBDB亦致力于此,它对于宋代文人的信息收录比较完备,相较之下明清则弱一些,这就留下了依据家谱进一步深入开拓的空间。

② 来集之《赠言小序》,(民国)《萧山来氏家谱》卷五《赠言》。

③ 《贵州参议继山来公暨配周宜人墓志铭》,陶望龄《歇庵集》卷八,台北:伟文图书出版社有限公司1976年版,第1137—1145页。

④ 《歇庵集》卷一〇,第1489—1494页。

⑤ 《歇庵集》卷九,第1401—1406页。

⑥ 《寰宇通志》称戴子静为至正二年进士,萧启庆以(万历)《绍兴府志·选举·进士》无其名,且同书《职官志·学职》称其为至正七年乡贡,认为其"应非进士"。详见《元代进士辑考》,台北:台湾"中央"研究院历史语言研究所2012年版,第306页。

⑦ 魏骥、钱福、潘士藻现存文集中并无家谱中所存之墓志铭、像赞、行状等。

础。除了这些带有自然属性、交往属性的数据之外,还可以从家族治经、家族文学等角度挖掘社会网络关系。

治经问题,一直是科举研究中的重点,它首先是一个经学选择、习得的问题;其次,就进士家族而言,还涉及家族内部的经学传承;第三,就地域而言,可能还存在着此地与彼地经学选择的差异。之前学界多关注“一经传家”“地域专经”等,也取得了一系列的优秀成果①。通过对哈佛大学提供的“包含明朝52 科的进士登科录”②共 14116 名进士信息的初步分析,发现明代萧山进士中治《诗经》者 3 人,治《易》者 2 人,治《书》者 24 人——这其中就包含了来天球、来汝贤、来经济与来三聘,这 4 人皆治《书》。在来氏家族身上,完美地体现出“一经传家”与“地域专经”的统一。当然,这种“统一”往往带有偶然性,除非我们能证明两者之间有必然的联系。

“一经传家”问题比较复杂,其中一个方面涉及本家族内部的经学习得、经学传承。当一个个地域相近(亦或同地)、时代相仿且又具有“一经传家”传统的进士家族以具有自己家族经学传承特色的姿态出现在科考舞台时,我们不禁要问:在经学传家方面这些进士家族是否有相似性,亦或各具特色?他们的传家之经是否一致?如果不一致,所传之经能否对其他家族的经学习得、传承产生影响?如果产生影响,可能会通过什么途径?进士家族之间的联姻,可能是解决这一问题很好的切入点。

上文中我们以来天球子女的姻亲情况为例,绘制出了来氏、邵氏、江氏与惠氏、徐氏、张氏几大家族的姻亲图谱(图 4),这六大家族中,来氏、邵氏、江氏可以被视为“进士家族”。就现有文献来看,明代的萧山来氏以《书》传家。那么,这个家族的“以《书》传家”的传统会影响到与其联姻的其他家族或者受到其他家族的影响吗?

先看余姚邵氏。余姚一地,早期多专《礼记》,后来专《易》者渐多。邵氏家族,情况要复杂一些,该家族有明一代共有 17 位进士,本经为《礼记》者 10 人、为《易》者 4 人、为《书》者 3 人③,基本上也与余姚当地的“地域专经”特点相吻合。

邵时顺娶来天球次女,邵漳为二人之子,生于正德十年(1515),此年距离其祖父邵蕃登成化二十年(1484)进士第已 32 年、距外祖父来天球中式也已 26 年,可以想见邵漳的成长与就学,可能同时受到本家与外家治学理念的影响。嘉靖二十三年(1544),邵漳以家传之《礼记》中式。万历八年(1580),邵漳从子邵梦弼也以《礼记》为本经中式,续写了家族专治《礼记》的传统。其实,从邵蕃到其孙邵漳,再到邵梦弼,他们延续的可能是邵氏家族道四房的传统。这一支中,早在永乐二十二年(1424),邵宏誉便以《礼记》为本经中式,成为该家族在明代的第一位进士。前贤时彦的深刻影响,早期家族内部较为稳定的《礼记》传统,使我们似乎看不出外家对于邵氏治经选择的丝毫影响。

再来看仁和江氏。明代仁和一地,以《易》中式者最多,《诗经》次之,《书》与《春秋》相仿佛,《礼记》最少。江氏五代进士,江玭、江澜父子以《礼记》中式。娶来天球三女的江晖,与祖、父二人专治《礼记》

① 如鹤成久章《明代余姚的〈礼记〉学与王守仁——关于阳明学成立的一个背景》(吴震、吾妻重二主编《思想与文献:日本学者宋明儒学研究》,上海:华东师范大学 2010 年版,第 356—367 页)、《明代安福の〈春秋〉学——举业から見た学問的系譜》(《福冈教育大学国语科研究论集》第 56 号,2015 年)、《明代莆田的〈书经〉学——福建的学术与举业关系的一个考察》(《教育与考试》2016 年第 1 期),陈时龙《明代科举之地域专经——以江西安福县的〈春秋〉经为例》(《台湾“中央”研究院历史语言所集刊》2014 年第 3 期,第 359—426 页)、《明代无锡的科举与〈尚书〉经》(《明清论丛》2016 年第 1 期,第 52—83 页)、《明代科举与地域专经》(《中国社会科学报》2017 年 8 月 22 日第 7 版)等。

② 实际该文件仅包含 51 科、14116 名进士信息。《天一阁藏明代科举录选刊·登科录》中收录的《嘉靖二年进士登科录》《崇祯十三年庚辰科进士三代履历》本文件并未收录。2017 年 10 月 1 日访问链接如下:http://projects.iq.harvard.edu/files/cbdb/files/ming_jinshilu_52y_release.xlsx? m=1438181509。

③ 参见拙文《大数据视阈中的明代登科录研究——以余姚进士家族为中心》,《清华大学学报》2019 年第 2 期。

不同,在正德十二年(1517)以《春秋》登进士第。如果从治经的选择来看,本家族的影响可能要远远大于来氏家族——江晖其父中式要比其岳父来天球早 13 年,且祖、父二人皆治《礼记》。江晖既没有选择《礼记》,也没有选择《书》,而是专治《春秋》——其从子圻、圻子铎皆以《春秋》中式,或许是深受江晖的影响。同样,在治经的选择问题上,江氏与来氏亦未发生交集。

通过以上梳理,我们并没有发现来氏对邵氏、对江氏两个家族在治经选择方面的影响,同样,也没有发现邵氏、江氏两个家族对来氏治经的影响。"没有发现"并不意味着这种影响不存在,可能只是囿于文献或思路问题暂时没有解决。姻亲家族之间在教育子孙后代方面,常有比较紧密的联系。例如清溪沈氏家族第十七世沈初,乾隆二十八年(1763)登进士第。31 年后,在《兰韵堂诗集·自序》中,沈初回忆自小依外祖居且从外祖就学之事,仍是颇多感慨:"余生三岁,从外祖识字即辨四声,稍长渐知五七言句法……今老矣,犹忆从外祖学,一长几横窗下,帘阴寂然,外祖时一讽咏,余辄心识之。"①冯建民曾以钱陈群之母、何士祁之母、李滋然之母、吴锡嶲之母等为例,说明家族中的女性在课子过程中也会承担教学责任,特别是对一些来自书香门第的女性而言②。我们再向前推进一步,如果这些女性恰好自本家族中习得某经、长于某经,进而以此经教授其子孙辈亦在情理之中。当然,现阶段这一问题并无实例可以佐证,权且提供一种思路而已。如果这一问题能在一定程度上获得解决,那么它所涉及的内容将有利于推动联姻的进士家族在治经选择等方面的研究,也有利于用社会网络分析来再现进士家族与教育、与治经等领域的关系。

明清时期,经学一直是科举考试最重要的内容,载体就是儒家经典,这使得一些家族在擅长经学的同时,往往也熟悉儒家典籍,有可能成为文学家族。

在对中华寻根网的家谱目录进行数据抓取时,发现不少家谱中保存着家族成员的诗文作品,有些还以家集的形式存在③。例如清溪沈氏,是我们利用爬虫程序从中华寻根网中发现的拥有家集的进士家族,潘光旦、龚肇智等都曾予以关注。

光绪十二年(1886)《清溪沈氏家乘》显示,沈文俊为该家族清溪始祖,三世沈珍五子,次子沈昊,是为碧澜公,《清溪沈氏家乘》所涉及的主要就是该支;四子沈昪,为东林公,另立世系,其子沈琮、沈琏分为正统七年(1442)、景泰二年(1451)年进士,此支暂且不论。碧澜公沈昊支在明清两代共有 9 位进士,明代 7 位,分别为六世沈棨,七世沈炼,八世沈圻、沈垔、沈垣,十一世沈萃桢、沈中柱;清代 2 位,为十三世沈崐、十七世沈初。

家谱中有文集两卷(卷十四、卷十五),共收入自五世沈渭至十八世沈应奎,共 29 人 110 篇作品,六世沈棨与十七世沈初数量最多,皆为 17 篇;诗集三卷(卷十六至卷十八),涵盖了自三世沈珍至十九世沈葆华共 92 人的 1431 首诗歌,沈棨、沈不负、沈修龄分别以 254 首、113 首和 68 首分列前三位;词、赋共一卷(卷十九),卷十九上收录自沈棨至沈应奎,共 16 人 152 首词,数量居前三位的分别是沈岸登(34 首)、沈修龄(26 首)和沈皞日(19 首);卷十九下收录自沈崐至十九世沈光翰,共 9 人 21 篇赋。

就词人与词作而言,16 位沈氏家族词人跨越明清两代。《全清词·顺康卷》收录了沈皞日、沈岸登、沈崐以及沈晭四人的部分作品,其他 12 位词人与词作,《全明词》《全明词补编》以及《全清词·顺康卷补编》《全清词·雍乾卷》等皆失收。

《全清词·顺康卷》所收的这 4 位词人,有家谱存词而《全清词》失收的情况。如沈皞日《燕山亭·七

① 沈初:《兰韵堂诗集·自序》,《兰韵堂诗集》,乾隆甲寅(1794)刻本。
② 冯建民:《清代科举与经学关系研究》,上海:华中师范大学出版社 2016 年版,第 246—247 页。
③ 例如萧山来氏,便有家集《冠山逸韵》《冠山逸韵续编》,现藏于国家图书馆。徐雁平《清代家集叙录》(合肥:安徽教育出版社 2017 年版)考订甚详,可参看。

夕饮华亭兄寓归而赋此以志昔怀》《喝马一枝花·用竹坨送余之来宾韵写意》《喝马一枝花》等三首,《全清词·顺康卷》及补编皆失收①。康熙二十三年(1684),朱彝尊与洪昇、严绳孙、彭孙遹、曹贞吉、朱昆田、沈岸登等送沈皞日之官来宾,朱彝尊作《一枝花·送沈融谷宰来宾》,一时题作云集,严、彭、曹、朱同题词作皆见于《曝书亭集》②。家谱中除了收有沈皞日次朱彝尊词韵的两首《喝马一枝花》外,还有沈岸登《喝马一枝花·送柘西叔之任来宾》,亦当作于同时。借助词作呈现出来的唱和活动,可以为我们进一步完善家族成员之间、家族成员与其他人士的交游等提供了详实可靠、且能通过第三方文献相印证的素材,这些正是深化社会网络分析的重要文献。还有文本相异的情况。《全清词》据《柘西精舍集》《瑶华集》所收沈皞日词与家谱所收,多有字句相异之处。再如沈皞日《摸鱼儿·莼》一首,《柘西精舍集》本下片与《瑶华集》本下片不同,家谱所收除上片个别字句外,基本与《柘西精舍集》本相一致。如果能将几种版本对勘,或许有新的收获。沈皞日是“浙西六家”之一,家谱所存之词为我们深入考察词作的生成、递改提供了范本。

其实,利用技术手段从家谱入手来考察家族成员的作品,甚或家集,整理文献是题中应有之义。同时,还可从以下领域深化研究:

第一,如果家谱中保存了足够多且有分量的本家族成员的诗、文、词、赋等作品,是否可以为某种文体单独作史,例如家族词史、家族诗歌史等。清溪沈氏,以词知名,在“浙西六家”中占据两席。该家族中有词作者 16 人,自三世至十九世,历明清两代,可否将沈皞日、沈岸登纳入到家族词学创作、发展与演变的长河中考察?可否将家族词学与明清两代的词风宗向对参?沈棨、沈崐皆是进士出身,除了词作外,还有诗、文存世,他们不同文体的作品在语源、表达等方面是否具有相似性?再如家族诗歌中有不少组诗,且很多与农事、村居相关,例如十二世沈不负、十三世沈岸登各有 13 首《农具诗》,十四世沈修龄有 30 首《村居诗》,沈金茎存诗 5 首皆为《村居杂吟》。擅长多与农事、村居相关的组诗,是否可以视为家族诗歌特色之一?如果是,这种特色有何师法、渊源?除了沈不负、沈岸登夫子自道向梅尧臣取法外,与当时的诗坛诗风宗向有无关系,这是否意味着该家族有学宋诗的传统?这些都可以深入探讨。

第二,书信、题跋、传略、墓志等作品中,往往包含着与他人交往的重要信息,有一些极富史料意义。例如沈中柱是刘宗周门人,《刘蕺山集》《明儒学案》中都有刘宗周答沈中柱书,恰巧家谱中保存了沈中柱《上刘念台老师书一》《上刘念台老师书二》,可以在一定程度上还原师生论学问道的场景。同时,也可以将沈中柱放到蕺山及门下弟子的论学交往圈中考察,扩大了社会网络分析的考察范围。

简单来看,引入家谱中的诗文词赋等作品或者家集,主要有三个意义:一是整理家族存世文献;二是从文学发展流变的角度来考察相应作品,将家族作品放到文学发展的长河中,放到地域文学的对照中进行纵向与横向关照;三是利用作品所反映出来的人与人之间、人与群体之间、家族与群体之间的复杂关系,建构社会网络谱系,将以血缘、姻亲为主的世系图,与社会交往属性更复杂的谱系结合起来,从而实现对进士家族的全面而整体的关照。

回到开篇那个问题,现阶段最直观判断某时段、某区域进士家族的方法是利用 CBDB,它在带来便利的同时,也暴露出匹配不准确、字段多重复等问题——对关系型数据库而言,这是比较严重的问题。

于是,我们便尝试利用爬虫程序从中华寻根网的家谱中直接提取信息,判断进士家族。为了梳理进士家族的繁衍过程,还引入了社会网络分析,借助技术手段将家族中的相关成员呈现在世系图中。世系图可以呈现血缘、姻亲两种关系,前者着眼于本家族的支脉传承,后者着眼于家族间的联姻,将之前一个

① 现阶段收录沈皞日词最为详备的当推胡愚、朱刚点校整理的《柘西精舍词》(上海:华东师范大学 2015 年版)。胡、朱二位将刻于康熙三十九年(1770)、现藏于上海图书馆的《柘西精舍诗余》与《浙西六家词》本《柘西精舍集》相较,共得沈皞日词 176 首。

② 系年据张宗友《朱彝尊年谱》,南京:凤凰出版社 2014 年版,第 311 页。

个独立繁衍的进士家族与其他家族联系起来——本文中,主要以进士家族间的联姻为着眼点,当然,也可以将标准放得更为宽泛一些。进士家族联姻之后,彼此家族的治经传统是否会影响对方家族?就我们所举的几个联姻的家族例子来看,似乎并不明显——当然也有可能是数据量不够大,缺乏典型性,但这也是关于专经、传经问题的新思路。再有,家谱中留存的诗文词赋等作品也是继续完善进士家族社会网络分析的最原始、最有价值的材料之一。借助这些作品,一是可以准确判断出哪一支成员更擅长哪种文体,从而梳理出整个家庭的文学发展脉络;二是可以将本家族的文学发展与其他家族的文学发展、与整个时代的文学发展相对照,从中探讨本家族文学发展的独特性,亦或是趋同性等;三是可以借助数据挖掘,将家族文学作品中的时间、地点、人物、事件等客观信息提取出来,在家族成员之间、家族成员与其他相关人物之间建立起联系,从而进一步完善以血缘、姻亲为核心的世系图谱,尽可能更全面地呈现出进士家族的社会网络分析图谱,这正是借助于大数据推动明清进士家族研究的意义之一。

A Study on the Families with Palace Graduates in the Ming and Qing Dynasties from the Perspective of Big Data

Liu Jingchen

(*Research Institute of Literature*, *Chinese Academy of Social Sciences*, *Beijing* 100732, *China*)

Abstract: As a relational database, China Biographical Database Project (CBDB, 中国历代人物传记数据库) has become the priority when the families with palace graduates (进士) in the Ming and Qing Dynasties is examined from the perspective of big data, since it containsover 420,000 items of data from biographies of historical figures of various dynasties. Its inquiry of "kinship relationship" can show the networks of kinship relationship, thus facilitating the sorting of the relations in the families with palace graduates. Meanwhile, it must be admitted that there exist such problems in CBDB as the inaccuracy in matching and the repeating of fields. These problems can be solved by checking such documents as archives of family genealogy. At present, www. ouroots. com.cn (中华寻根网) provides the images of the full texts of 2,392 archives of family genealogy, enabling us to pick up the related information in the catalogs of the archives with the help of the crawler program (爬虫程序), which canautomatically sort out families with palace graduates and offer a favorable supplement to CBDB. On this basis, and by way of the analysis of social networks, maps of genealogies of families with palace graduates centering on relationship by blood and relationship by marriage can be clearly presented.Relationship by marriage could bind together different families with palace graduates, thus broadening our view from a single family to several families with palace graduates that were related to each other. For example, we can observe if there existed certain influence among these families with relationship by marriage when they studied and disseminated Confucianclassics. The literary works of family members that were preserved by archives of family genealogy laid down the foundation for both observing the development and evolution of family literature, and constructing the maps to analyze more extensive social networks.

Key words: big data, families with palace graduates(进士) in the Ming and Qing Dynasties, CBDB (China Biographical Database Project, 中国历代人物传记数据库), www.ouroots.com.cn(中华寻根网), family genealogy, analysis of social networks

(责任编辑　　管　琴)